

The Delftia Project: Engaging Citizen Scientists in Collaborative Knowledge Creation

Lauren Ramilo, Daiza Norman, Rabeya Tahir, Tanasha Lertjanyarak, Carlos Goller

Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695




Introduction

- High-throughput and metagenomic sequencing have generated tremendous amounts of data regarding environmental microbes. Multidisciplinary interest and research efforts have produced a complex and disordered body of literature, hindering inreach and outreach communications.
- To align research goals and encourage public dissemination, we must improve access, mining, discovery, reuse, and interoperability of literature.
- We propose the use of collaborative annotation driven by citizen scientists, which creates a conversational layer above publications to host group discussion and knowledge transfer. Information extraction and evaluation is expected to improve accessibility and inform communication.

Hypotheses

1. Collaborative annotation assists the comprehension of published literature
2. Annotations can be used to produce both narrative and logical-scientific content.

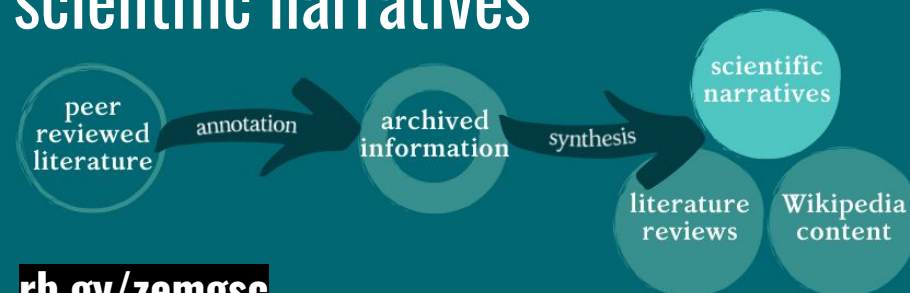
Methods

 1 Alchemy Blog <ul style="list-style-type: none">• Engage, train, and inform annotators• Communicate topics• Alt text and SEO for accessibility	 2 Collaborative Annotation <ul style="list-style-type: none">• In-browser annotation with Hypothes.is• Retrieved via CROWDLAERS and jonudell.info	 3 Resource creation <ul style="list-style-type: none">• Narratives on the <i>Alchemy Blog</i>• <i>Delftia</i> spp. on Wikipedia• Literature review
---	---	---

Results

- Flesch-Kincaid grades for annotations and text are 9.6 and 12.9, respectively.
- Annotations added content (55%), questions (29%), summary (9%), emotion (8%), or other (3%) (n=415).
- Multimedia is present in 9% of annotations.
- Discussion yielded 17 threads with a maximum depth of 2 replies.
- Half (51%) of annotations were not tagged. Tagged annotations often used >1 tags (63%, n=116).
- Annotations informed the creation of 3 narratives and 7 logical-scientific resources.
- Logical-scientific content ($\bar{x}=19 \pm 19.6$) uses more references than narratives ($\bar{x}=14.3 \pm 12.0$), although there is considerable deviation

Collaborative annotation of research literature by citizen scientists improves comprehension and informs scientific narratives



rb.gy/zemgsc



SCAN ME

	Total	Average Annotations
Articles	98	4.78 ±5.04
Days	193	2.41 ±4.99
Users	14	33.43 ±61.10

Table 1 Summary of Annotation Activity
98 articles were annotated (n=43) over 193 days (x=2.4) by 14 users (1-33-4). Average annotations per user has a high standard deviation (4.78), indicating large variations in activity between users. (n_{user}=268)

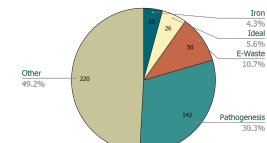


Figure 1 Annotation of Specific Topics
Annotations focused on specific topics such as interactions with iron (n=20), Ideal Environment for Gold Precipitation (n=26), Pathogenesis (n=142), and Electronics Waste (n=20). Annotations labelled Other (n=220) include glossing, general information, and annotations used to inform Wikipedia pages. (n_{user}=268)

Conclusions

- Suggested topics communicated through the *Alchemy Blog* were annotated sufficiently, indicating a strong focus. Annotations categorized as “other” do not align with topics but are still valuable, containing glossing, general information, or information used for Wikipedia (Figure 1).
- Annotations improved the readability of scientific texts by explaining concepts in simple terms, or supplementing with context and multimedia. Information sharing, as well as questions and discussion, enhances collaborative learning and comprehension of the text.
- Retrieving relevant annotations is easiest via an organized tag system. However, because only half of annotations were tagged, annotation and source text were searched instead. We did not establish guidelines and created tags as needed. This likely contributed to poor tag use.
- Collaborative annotation successfully enabled the creation of both logical-scientific communication and narratives. Narratives are unique from their use of the traditional storytelling format, and are often more engaging than other forms of science communication. They can be improved even further when a wealth of archival information is weaved into the story, producing engaging and informative content.

In the future

Continued annotation will contribute to a literature review and create more content for topics of interest. We anticipate that this will promote interest and catalyze research into *Delftia* and it's potential applications.

References

- Atrash A, Abel MH, Moulin C. 2014. Supporting organizational learning with collaborative annotation. Proceedings of the 6th International Conference on Knowledge Management and Information Sharing, Rome, Italy. p. 237-244.
- Li X, Nie Y, Chen M, Liu X, Liu X. 2014. Member's behavior in virtual learning community: A study using data mining approach. Comput. Inf. Sci. 7(4).
- Glover I, Xu Z, Hardaker G. 2007. Online annotation -- Research and practices. Comput. Educ. 49:1308-1320.

Acknowledgements and Contacts

We thank the members of the *Delftia* Hypothes.is group for all of their contributions, support, and for lending some mental processing power. We thank Hypothes.is for their support and advice. We thank Remi Kalir for his insights, pedagogical knowledge, and CROWDLAERS. We thank NC State University Open Pedagogy group for their inspiration and support, and the Office of Undergraduate Research for funding this project. Special thanks to Nia and Kotchka. We are endlessly grateful for everyone who contributed to this project, which wouldn't be possible without your support and knowledge.



NC STATE UNIVERSITY